

Session 6:

DATA INGESTION TOOL SQOOP &
INTRODUCTION TO CASE STUDY 1

Assignment 1

TASK – 1:

Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

EXECUTION:

Data is residing in mysql database table "Person",

```
File Edit View Search Terminal Help
ERROR 1062 (23000): Duplicate entry '6' for key 'PRIMARY'
mysql> insert into Person values (7,'Akash','Gupta','Andheri','Mumbai');
Query OK, 1 row affected (0.00 sec)

mysql> insert into Person values (8,'Isha','Kapoor','Andheri','Mumbai');
Query OK, 1 row affected (0.00 sec)

mysql> select * from Person;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
| 1 | Shyam | Ram | Patna | Bihar |
| 2 | Tanya | Priya | Whitefiled | Bangalore |
| 3 | James | Brown | New York | United States |
| 4 | Jhon | Miller | Los Angeles | United States |
| 5 | Harry | Potter | Inner Harbor | Baltimore |
| 6 | Vijay | Kumar | Anna Nagar | Chennai |
| 7 | Akash | Gupta | Andheri | Mumbai |
| 8 | Isha | Kapoor | Andheri | Mumbai |
| 789 | a | b | c | d |
+-----+-----+-----+-----+-----+
9 rows in set (0.00 sec)

mysql>
```

The data has to be imported in to HDFS directory using SQOOP command,

```
sqoop import --connect jdbc:mysql://localhost/simplidb --username root --password Root@123
--table Person --m1 --target-dir /sqoopout/Person
```

After the running the sqoop command, the job executed successfully and the records are imported in to HDFS folder path,

```
acadgild@localhost:~/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/
File Edit View Search Terminal Help
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=15619
Total vcore-milliseconds taken by all map tasks=15619
Total megabyte-milliseconds taken by all map tasks=15993856
Map-Reduce Framework
  Map input records=9
  Map output records=9
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=281
  CPU time spent (ms)=4460
  Physical memory (bytes) snapshot=108908544
  Virtual memory (bytes) snapshot=2063437824
  Total committed heap usage (bytes)=32571392
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=279
19/03/15 22:16:40 INFO mapreduce.ImportJobBase: Transferred 279 bytes in 56.6267
seconds (4.927 bytes/sec)
19/03/15 22:16:40 INFO mapreduce.ImportJobBase: Retrieved 9 records.
[acadgild@localhost bin]$
```

Querying the target HDFS folder path shows the data is imported,

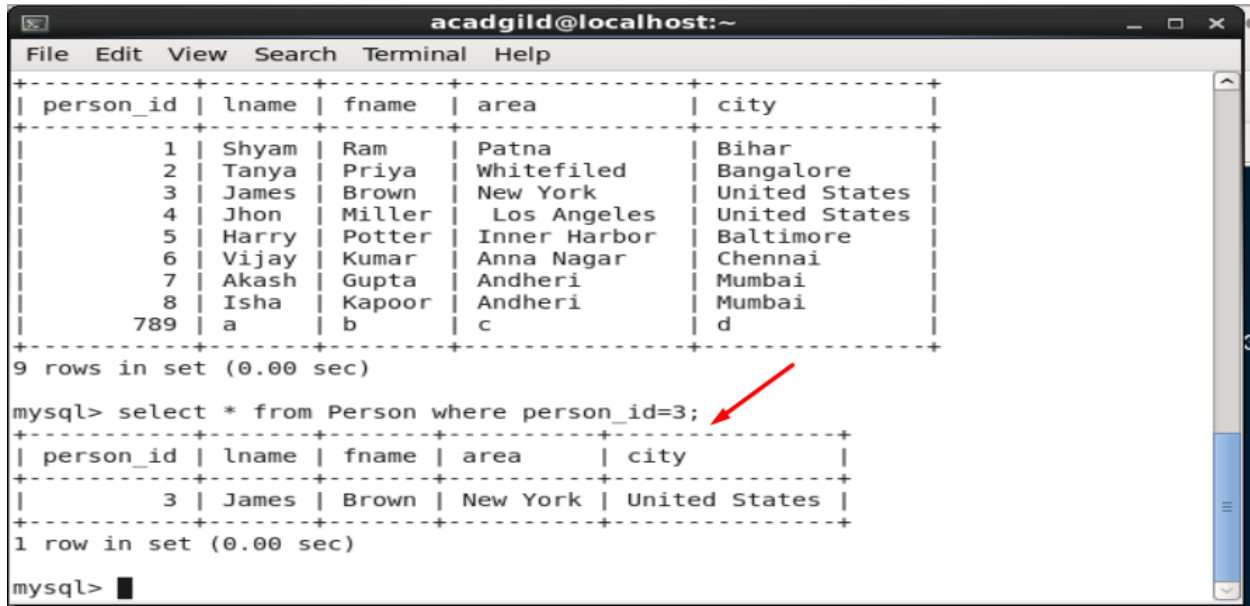
```
acadgild@localhost:~/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/
File Edit View Search Terminal Help
19/03/15 22:16:40 INFO mapreduce.ImportJobBase: Transferred 279 bytes in 56.6267
seconds (4.927 bytes/sec)
19/03/15 22:16:40 INFO mapreduce.ImportJobBase: Retrieved 9 records.
[acadgild@localhost bin]$ hadoop fs -ls /sqoopout/Person
19/03/15 22:24:30 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2019-03-15 22:16 /sqoopout/Person/
_SUCCESS
-rw-r--r--  1 acadgild supergroup      279 2019-03-15 22:16 /sqoopout/Person/
part-m-00000
[acadgild@localhost bin]$ hadoop fs -cat /sqoopout/Person/part-m-00000
19/03/15 22:24:59 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
1,Shyam,Ram,Patna,Bihar
2,Tanya,Priya,Whitefiled,Bangalore
3,James,Brown,New York,United States
4,Jhon,Miller, Los Angeles ,United States
5,Harry,Potter,Inner Harbor,Baltimore
6,Vijay,Kumar,Anna Nagar,Chennai
7,Akash,Gupta,Andheri,Mumbai
8,Isha,Kapoor,Andheri,Mumbai
789,a,b,c,d
[acadgild@localhost bin]$
```

TASK – 2:

Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table with parameter person_id =3.

EXECUTION:

As per the task, we need to import person record whose person id is 3 from mysql database table "Person" as below,



The screenshot shows a terminal window titled 'acadgild@localhost:~'. It displays a MySQL query and its results. The query is 'select * from Person where person_id=3;'. The results show a single row with person_id 3, lname James, fname Brown, area New York, and city United States. A red arrow points to the query text.

```
mysql> select * from Person where person_id=3;
```

person_id	lname	fname	area	city
3	James	Brown	New York	United States

```
mysql>
```

The below sqoop command should be executed to filter the records and import it to HDFS target folder path,

```
sqoop import --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person --m1 --target-dir /sqoopout/FilteredPerson --where "person_id=3"
```

Once the command is executed successfully, query HDFS folder path for the resultant file,

```
acadgild@localhost:~/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/ _ □ x
File Edit View Search Terminal Help
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=6489
Total vcore-milliseconds taken by all map tasks=6489
Total megabyte-milliseconds taken by all map tasks=6644736
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=118
  CPU time spent (ms)=1820
  Physical memory (bytes) snapshot=107999232
  Virtual memory (bytes) snapshot=2063437824
  Total committed heap usage (bytes)=32571392
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=37
19/03/15 22:42:53 INFO mapreduce.ImportJobBase: Transferred 37 bytes in 34.7447
seconds (1.0649 bytes/sec)
19/03/15 22:42:53 INFO mapreduce.ImportJobBase: Retrieved 1 records.
[acadgild@localhost bin]$
```

```
Bytes Written=37
19/03/15 22:42:53 INFO mapreduce.ImportJobBase: Transferred 37 bytes in 34.7447
seconds (1.0649 bytes/sec)
19/03/15 22:42:53 INFO mapreduce.ImportJobBase: Retrieved 1 records.
[acadgild@localhost bin]$ hadoop fs -ls /sqoopout/FilteredPerson
19/03/15 22:44:25 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2019-03-15 22:42 /sqoopout/Filtere
dPerson/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 37 2019-03-15 22:42 /sqoopout/Filtere
dPerson/part-m-00000
[acadgild@localhost bin]$ hadoop fs -cat /sqoopout/FilteredPerson/part-m-00000
19/03/15 22:45:03 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
3,James,Brown,New York,United States
[acadgild@localhost bin]$
```